METU Studies in Development, 35 (December), 2008, 427-444

# A comparison of JMP wage decomposition and quantile regression methods in wage inequality assessment

Nazmi Yağanoğlu
*Çanakkale Onsekiz Mart University, Department of Economics, Çanakkale, Turkey*
*e-mail: nazmiyagan@lycos.com*

Hakan Ercan
*Middle East Technical University, Department of Economics, Ankara, Turkey*
*e-mail: hercan@metu.edu.tr*

**Abstract**

The decomposition technique of Juhn, Murphy and Pierce (1993) and quantile regression are two of the main tools of wage inequality analysis. JMP technique has the advantage of decomposing the change in wages into three components, and showing residual inequality easily. Quantile regression has the advantage of showing a detailed picture of wage distribution at different quantiles. We apply both techniques to March Current Population Survey (CPS) data of the US Bureau of Labor Statistics (BLS) to analyze the changes in wage inequality in the US during the 1967-2005 period. We compare the results to see which technique produces more useful results in response to the research question at hand. We find that it is a good idea to check the quantile regression results before concluding on JMP values since if quantile regression coefficients are very different from OLS coefficients (meaning the wage distribution is quite different from a normal one), results of two methods differ greatly and the application of JMP is problematic.

*Keywords*:Wage inequality, US, wage decomposition, quantile regression.

*JEL codes*: J31, C14.

## 1. Introduction

There has been a large literature, which documents the increase in wage inequality in the US during the 1980's (see for example Katz and Murphy, 1992; Levy and Murnane 1992; Bound and Johnson, 1992). Initially, these studies adopted a wage decomposition method that was suggested and used by Juhn, Murphy, and Pierce (1993) (JMP from this point onwards). This method became quite popular. Later, quantile regression methods, among others, were applied to the analysis of trends in wage inequality. This paper takes stock by applying both methods to the same data set and compares the outcomes.

JMP's decomposition method has been used in a number of cornerstone articles as well as their groundbreaking papers (1991 and 1993) and , for example Blau and Kahn's (1994 and 1996) analysis of gender and international comparisons of wage inequality. It has also been used to analyze the wage inequality in many other countries as well (for Russia, Brainerd (1998); for Estonia and Slovenia, Orazem and Vodopivec (2000); for India, Kijima (2005) among others). This method's advantage is that it decomposes the change in wage inequality into three separate effects of observable prices, observable quantities and unobservable prices and quantities. This is an important contribution in the sense that it shows the share of "within" inequality in overall wage inequality. Within inequality is simply the wage inequality among workers with similar observable characteristics and quite hard to identify because of the unknown nature of its causes. Sometimes it is called "residual inequality" as well, due to the way it is represented by the residuals of a wage regression.

JMP's decomposition technique has been criticized for the sensitivity of its results to the order of decomposition (Autor and Katz, 1999). Another criticism brought against it is about the way the changes in the distribution of residuals are modeled, which sometimes makes the total of decomposed components to be slightly different from the observed total change. (Lemieux, 2002).

Some later studies used quantile regression techniques (see, for example, Buchinsky, 1994 and 1998; Machado and Mata, 2001; Martins and Pereira, 2004). This technique is quite handy since it shows the effects of covariates on the distribution of the dependent variable at different quantiles. Running a number of quantile regressions, one can obtain a very useful tool of observing the relative importance of covariates at different parts of the wage distribution. Although the JMP method also compares the conditional quantiles of wage distribution, their technique is based upon ordinary least squares estimation, and thus the conditional mean. Naturally, they assign the same coefficients for the covariates for any point on the distribution of wages. We show how their estimates compare to those of quantile regression in Section 3.

Using March Current Population Survey (CPS) data of the US Bureau of Labor Statistics (BLS), we find that the wage inequality in the US, which increased remarkably during the 80s, has been still increasing at a slower rate. Although most of this increase came from the lower half of the wage distribution until the mid-80s, the change in the later years came almost exclusively from the dispersion in the upper half of the wage distribution. This is typically the sort of change that JMP method was designed to capture. In this work, we use the quantile regression method to generate the conditional quantiles equivalent to those created by the JMP method and compare them. We find that they might report similar values if the conditional wage distribution does not change much at different quantiles. However, there might be huge differences otherwise.

The rest of the paper is organized as follows: In section two, we briefly describe the JMP wage decomposition and quantile regression models that we use here. Then we describe the data set and variables. In section three, we present, compare and contrast the results. In section four, we conclude by stating the lessons that we have learned from this exercise.

## 2. Methodology and data

### *2.1. Data*

In this study, we have used the Current Population Survey (CPS) March Annual Social and Economics Supplement (ASEC) for years 1967 to 2005. The CPS data is collected monthly by interviewing a large number of households. Each household is interviewed once a month for four months in a row every year, then interviewed again next year during the same four months.

The main purpose of the survey is to collect employment information within the United States of America. However, it also contains data on demographic characteristics of the population such as age, race, number of children, area of living etc. The data is organized to give three perspectives: household, family and person. CPS is conducted among the civilian, non-institutional population in every state of the United States of America and the District of Columbia. ASEC is released for the month of March every year, including everything the other months' data have as well as additional information on work experience, income and migration. The related raw data files concerning can be downloaded from National Bureau of Economic Research (NBER from this point on) web site. The data in this web site is named in a somewhat misleading way. Because of the way the survey is conducted, a given year's data is published during the next year, and labeled the same as the year of publication. That is, the data labeled as 2006 in the above link actually belongs to year 2005. Throughout our analysis, whenever we mention years, we mean the real year that the data belongs.

We have limited our analysis to the males who are between the ages of 16-64, and work full-time and full-year (defined as working at least 40 hours a week and 35 weeks a year). They must have earned at least $67 in 1982 dollars per week (half of the real minimum wage based on a 40-hour week in 1982 dollars), have at least 1 year of potential labor market experience (defined as: age-years of education-6), not living in group quarters, not self-employed or working without pay.

We will use the log values of weekly wage and salary income throughout our studies since we make comparisons of different quantiles and years, and log values have well-known conveniences in this sort of comparison.

We have used four education dummies in our regressions: Less than high school (lowedu-less than 12 years of education), high school graduate (HS-12 years of education), some college (scedu-between 12 and 16 years of education) and college graduate (colledu-16 years or higher)[1].

### *2.2. JMP Method*

First applied by JMP, this method decomposes the changes in wages into the effects of changes in observable individual characteristics (education, experience etc.), the effects of changing skill prices of these observable skills, and changes in the distribution of residuals, using a wage equation and comparing for different percentiles. It starts with a wage equation such as:

$$Y_{it} = X_{it}\beta_t + u_{it}$$

where $Y_{it}$ is the log weekly wage for individual i in year t, $X_{it}$ is a vector of individual characteristics and $u_{it}$ is the part of wages accounted for by the unobservable characteristics, defined by an individual's percentile in the residual distribution, $\theta_{it}$, and the distribution function of the wage equation residuals $F_t$( .). These residuals can be expressed as:

$$u_{it} = F_t^{-1}(\theta_{it} \mid X_{it})$$

The wage equation can be manipulated so as to capture three sources of inequality: changes in the distribution of X's, changes in $\beta$'s and changes in $F_t$( .)

$$Y_{it} = X_{it}\bar{\beta} + X_{it}(\beta_t - \bar{\beta}) + \bar{F}^{-1}(\theta_{it}|X_{it}) + \left[F^{-1}(\theta_{it}|X_{it}) - \bar{F}^{-1}(\theta_{it}|X_{it})\right]$$

The first term on the right-hand side captures the effect of changes in observable characteristics, the second one captures the effect of changing skill prices of these observable characteristics, and the last one captures the effect of changes in the distribution of wage residuals. Then, one can restrict this last equation to find different wage distributions. For example, with fixed observable prices and fixed residual distribution, we have

$$Y_{it} = X_{it}\bar{\beta} + \bar{F}^{-1}(\theta_{it}|X_{it}). \quad (1)$$

[1] Other details on the data and a table of descriptive statistics are given in the Appendix.

Equation 1 attributes the change in wage distribution from one year to the other only to the changes in X's, observable characteristics. On the other hand, if we allow the quantities and prices of the observable characteristics change over time, the wage distribution can be generated by

$$Y_{it}^2 = X_{it}\,\beta_t + \overline{F}^{-1}(\theta_{it}|X_{it}) \tag{2}$$

Finally, allowing observable prices, quantities and the distribution of residuals to change in time, we have

$$Y_{it}^3 = X_{it}\,\beta_t + F^{-1}(\theta_{it}|X_{it}) = X_{it}\,\beta_t + u_{it} \tag{3}$$

When comparing two years in terms of inequality, the difference from one year to the other in (1) is attributed to the change in observable characteristics. The difference of this change from the change in (2) is attributed to the change in the coefficients of observable characteristics (skill prices); and the difference between the change in (3) and the change in (2) is attributed to changes in unobservable characteristics.

## 3. Quantile regression

In a wage equation model, we can define the quantile regression (Koenker and Bassett (1978)) setup as:

$w_i = x_i'\beta_\theta + u_{\theta i}$ , with $Q_\theta(w_i|x_i) = x_i'\beta_\theta$ ,

where $x_i$ is a vector of personal characteristics, $\beta_\theta$ is a vector of parameters and $Q_\theta(w_i|x_i)$ denotes the $\theta^{th}$ conditional quantile of $w$ given $x$. Any given quantile θ can be derived by solving the following problem:

$$\min_\beta \left\{ \sum_{\ln w_i \geq x_i\beta} \theta\,|w_i - x_i'\beta_\theta| + \sum_{\ln w_i \leq x_i\beta} (1-\theta)\,|w_i - x_i'\beta_\theta| \right\},$$

which can be written as

$$\min_\beta \sum_i \rho_\theta(w_i - x_i'\beta_\theta),$$

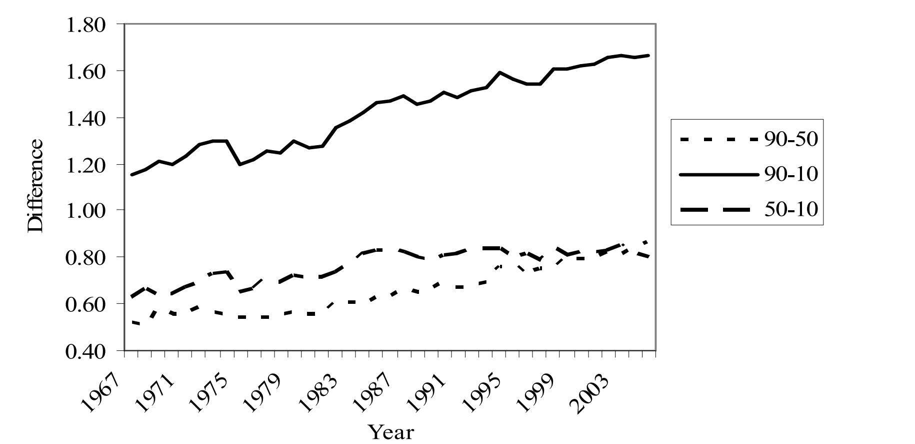
where $\rho_\theta(\varepsilon)$ is the check function defined as $\rho_\theta(\varepsilon) = \theta\varepsilon$ if $\varepsilon \geq 0$ or $\rho_\theta(\varepsilon) = (\theta-1)\varepsilon$ if $\varepsilon < 0$. Since the objective function is not differentiable, it is not possible to use standard optimization methods. It can be solved using linear programming methods. Generalized Method of Moments estimation is also possible (Buchinsky (1998)).

Increasing $\theta$ from 0 to 1, one can trace the whole distribution of $w$ conditional on $x$, taking snapshots along the way. The coefficient estimates of quantile regression denote the effects of covariates on the distribution of the regressor at the corresponding quantile, thus giving the user a means to compare distributions[2]. As Koenker and Bassett (1978) proved it, even though the estimator $\hat{\beta}_\theta$ lacks in efficiency compared to the least squares estimator in case of a Gaussian distribution, it is much more efficient and robust for a large array of non-Gaussian situations. Especially for the cases when the conditional distribution of the dependent variable (conditional on covariates) in question has thick tails, is asymmetric, or unimodal, the meaning attributed to the linear regression estimator can be made much stronger with the help of quantile regression estimators that provide better information about the distribution of the variable in question.

## 4. Wage inequality, JMP and quantile regression

Figure 1 gives us a picture of the change in overall wage inequality in the US during the 1967-2005 period. All the values reported are 3-year moving averages centered on the indicated year. We notice that the 90th-10th percentile difference of log weekly wages increases sharply for most of the 80s then keeps increasing at a slower rate. It shows that the $50^{th}$-$10^{th}$ percentile difference, representing the inequality in the lower half of the wage distribution, also increases until mid-80's, then stabilizes for the rest of the period. On the other hand, the $90^{th}$-$50^{th}$ percentile difference, which stands for wage inequality in the upper half of the wage distribution, has been increasing since the mid-1980s. The recent increase in wage inequality obviously comes from the change in dispersion in the upper half of the wage distribution.

***Figure 1 Wage Inequality, 1967-2005***




2 Researchers use this technique more recently. See Machado and Mata (2005) for a very promising setup.

The standard application of JMP is by decomposing the changes in one of the traditional measures of wage inequality: quantile differences. Table 1 shows JMP decomposition results for our sample in regular interval of years[3].

**Table 1**

Observable and Unobservable Components of Change in Inequality

| Year | Percentiles | Total Change | Components | | |
|---|---|---|---|---|---|
| | | | Obs.Quant. | Obs.Price | Unobservable |
| A.1967-1975 | 90th-10th | 0.03 | 0.06 | -0.04 | 0.01 |
| | 90th-50th | 0.01 | 0.02 | -0.01 | 0.00 |
| | 50th-10th | 0.01 | 0.04 | -0.03 | 0.01 |
| B.1975-1980 | 90th-10th | 0.04 | 0.08 | -0.13 | 0.09 |
| | 90th-50th | 0.01 | 0.01 | -0.06 | 0.05 |
| | 50th-10th | 0.04 | 0.06 | -0.07 | 0.04 |
| C.1980-1985 | 90th-10th | 0.17 | 0.03 | 0.06 | 0.08 |
| | 90th-50th | 0.06 | 0.00 | 0.03 | 0.03 |
| | 50th-10th | 0.11 | 0.03 | 0.03 | 0.05 |
| D.1985-1990 | 90th-10th | 0.02 | -0.01 | 0.01 | 0.03 |
| | 90th-50th | 0.06 | 0.02 | 0.02 | 0.02 |
| | 50th-10th | -0.04 | -0.03 | -0.01 | 0.00 |
| E.1990-1995 | 90th-10th | 0.09 | 0.02 | 0.02 | 0.05 |
| | 90th-50th | 0.07 | 0.03 | 0.01 | 0.03 |
| | 50th-10th | 0.03 | -0.01 | 0.01 | 0.02 |
| F.1995-2000 | 90th-10th | 0.06 | 0.00 | 0.03 | 0.03 |
| | 90th-50th | 0.05 | 0.01 | 0.02 | 0.02 |
| | 50th-10th | 0.01 | -0.01 | 0.01 | 0.01 |
| G.2000-2005 | 90th-10th | 0.04 | 0.02 | 0.01 | 0.01 |
| | 90th-50th | 0.04 | 0.04 | 0.00 | 0.00 |
| | 50th-10th | 0.00 | -0.02 | 0.01 | 0.01 |

Results reported are the 3-year averages centered around the indicated year, with the exception of 1967 and 2005, which are 2-year averages

[3] These results are obtained from application of JMP on the regression of log weekly wages on a quartic of experience, education dummies for less than high school, high school graduate, some college and college graduate, industry dummies and demographic dummies like married, white, metropolitan area and living in the south. All regressions are 3-year pooled regressions centered on the indicated year except 2005, which is a 2-year pooled regression of 2004 and 2005. Dropped education category is less than high school, dropped industry is agriculture.

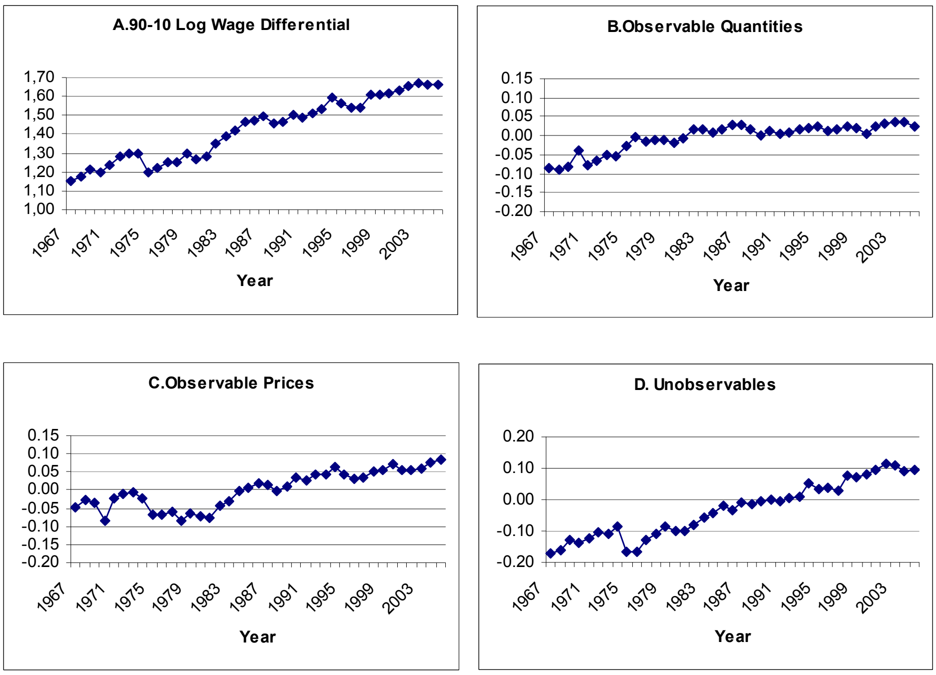
The main focus of the JMP technique is the importance of unobservable characteristics and their prices on wage inequality, since these represent the portion of wage inequality that cannot be explained by the characteristics that can be observed easily. One can clearly see from Table 1 that the effect of unobservable component is quite sizable compared to the total of observed components. This was put forward by the JMP as evidence of the importance of “within” or “residual” inequality, which is simply the wage inequality among workers with seemingly similar observable characteristics. The results we see above also show us that the effect of observable quantities on the overall inequality (90-10 difference) is limited compared to observable prices and unobservable component. It is only normal to expect that the skill structure in the economy cannot change too fast over time, while sometimes the skill prices might have to compensate for this lack of speed, depending on whether a particular skill is sought. We also see in Table 1 that in some cases observable prices and quantities work in different directions, cancelling the effect of each other in comparison to the unobservable part. This, of course, could be attributed to simple supply-demand mechanics at work for observable skills. However, the important thing about observable quantities and prices canceling each other’s effect is that it emphasizes the importance of unobservables on the changes in wage inequality.

We also see from Table 1 that our earlier analysis concerning upper versus lower half inequality is confirmed by the JMP results. Starting from 1985-90 period, the change in the upper half surpasses the change in the lower one. The separate effects for the three components also reflect the same trend in those years, with the exception of 2000-2005 period when an increase in upper half inequality and a decrease in lower half inequality for observable quantities offsets little movements in the opposite direction for observable prices and unobservable quantities and prices. The reinstatement of the effect of changes in observable quantities on the wage inequality could mean the start of a new trend with the new century, but it is too early to comment on that.

Another way of looking at the relative effects of these three components is by Figure 2. Panel A shows the observed change in 90-10 wage inequality for each year. Panels B, C and D show the decomposition of change in 90th-10th percentile difference for each year compared to the values obtained from a pooled regression. Thus, they are reported as difference from their long-term means. Naturally panels B, C and D do not add up to Panel A, since they are obtained from the regression results, while Panel A is obtained directly from the data. This structure lets us see if a component contributes to wage inequality significantly or not. The farther are the values from 0, the more contribution that component makes to the overall wage inequality for the given period of time.

We see from this picture that while the change coming from observable quantities is very limited and around zero especially after mid-1970s, the other two components differ from their long-term means. Among the latter two, the unobservable component seems to be effective throughout our period analysis, while observable prices (or skill premium) gain some importance only by mid-80s. Again we see that observable quantities seem to depart from their long-term average after the turn of the century, while the unobservables stabilize around their level. The trend is not very clear on observable prices in this period. These changes are not pronounced as strongly here as they were in Table 1, because the panels in Figure 2 are based on a decomposition of difference from long-term means.

**Figure 2**
The 90th-10th Percentile Change in Log Wage and Its Decomposition




On the other hand, the quantile regression focuses on taking pictures of wage distribution at different quantiles of wages. While the OLS estimators simply show the effect of the covariates on the regressor at the conditional mean, the quantile regression ones give the effects of covariates on the specified quantile of the *distribution* of the regressor. Panel A of Table 2 gives us a good example of this[4]. Here, we are able to compare OLS

4 The regressions were run on the same model as the JMP case. Reported here are the 3-year pooled quantile regression results.

estimates to quantile regression estimates at 5 different points in the wage distribution. We note that even though in some cases the OLS coefficient is not much different from the quantile regression coefficients in different quantiles (see for example high school 1995), many times we see a considerable diversion, meaning that there is a lot of untold story to be learned from the quantile regression coefficients. It is obvious that, for example, one cannot claim having a college degree meant the same thing at the 10th and 90th percentile in 1995.

The impact upon dispersion values reported in Panel B simply show the difference between the quantile regression coefficients for two percentiles. They also reveal how the level of education affects the difference between the wage distributions of those two quantiles. One can easily see that the main component of increasing wage inequality is college education. OLS coefficients support this result, too. For one thing, having higher education is a good thing. This is obvious from the fact that higher education brings in higher returns in all quantiles of the table. In addition, for the college-educated people, having a college education is much more important at the $90^{th}$ percentile than it is at the $10^{th}$ percentile after the 80s, indicating some within inequality among college people as well.

Using the quantile regression coefficients, one can also make comparisons over the years between coefficients of the same quantile. This information is not available from the OLS coefficients. For example, while the OLS coefficients for college degree keep increasing after 1975, those for the 10th percentile do not show such a clear trend. One can easily confirm that the increasing value of college degree that we see from OLS coefficients results mainly from its value for the higher percentiles, especially the 90th.

Although they are informative to the trained eye, the numbers in Table 2 are too many to be helpful to see the big picture. Figure 3 lets us have a look at the values in Table 2 from a different perspective. Each of its three panels shows how marginal effects for a certain level of education change over time. In Panel A we have the high school graduates; Panel B shows the same information for employees with less than three years of college and Panel C shows the values for people with at least four years of college education. As we have explained before, the OLS line acts as some sort of middle path here. We notice some trends in the behavior of the values of different quantiles compared to each other and OLS. For one thing, there is an obvious difference between college graduates and the other two groups. Up to mid-90s, the marginal effects are more or less similar between different quantiles for college graduates. On the other hand, the 80s are when the spread between different quantiles is the biggest for people with less than college degree, especially mid-80s. Then, the patterns switch.

Starting from the 90s, marginal effect of having a college degree starts to spread across the distribution, it gets narrower for the other two groups.

Another interesting movement can be seen when we compare the behavior of the lower tail to that of the upper tail. While the 10th percentile had the highest marginal effects of education for high school and some college groups until mid-80s, they saw sharp drops after that, ending up at the bottom in 2005. Even though the $25^{th}$ percentile followed from some distance in each case, there was not such a sharp drop there. For college graduates, with the exception of mid-80s, the lower half sees lower marginal returns all the way. At the other end of the distribution, we see that from mid-70s on there has been an increase in the marginal effects of education for people at the upper tail, no matter what level education they have.

This is a lot of information that is not going to be extracted from a single regression line. A natural question that comes to mind just looking at the JMP decomposition results in Table 1 is how it would look like if we calculated the same values using quantile regression coefficients. After all, part of what JMP report is the $90^{th}$-$10^{th}$ percentile difference of the conditional wage distribution. Using the average values of covariates for each year to predict wages for each quantile, we have done just that. The results are shown in Table 3 together with the JMP results, for comparison purposes.

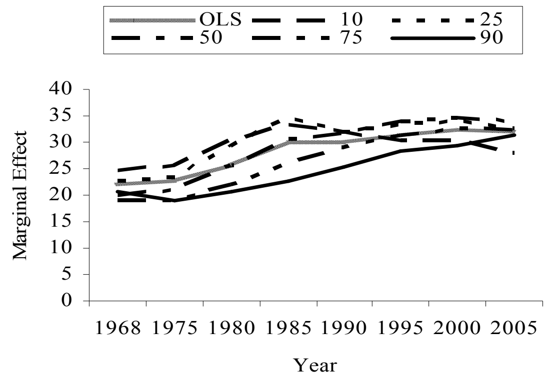
Interestingly enough, both methods report more or less the same results for the change from 1995 to 2000. The numbers for 1990-95 are reasonably close as well. On the other hand, they report totally different stories for 2000-2005 and 1985-90. Why is this so? The answer is in Table 2. We see that the change in coefficients for each quantile is similar between 1995 and 2000. This can also be claimed for 1990 and 1995. In addition, we notice that the quantile regression coefficients for these years are mostly very close to the OLS coefficient, indicating that the distribution has some symmetry. This is especially true for the two lower education groups. The college coefficients start to diverge after 1990, gaining speed at 2000 and 2005. However, we can say in general that quantile regression and JMP results in Table 2 are close when the quantile regression coefficients are close to the OLS coefficients.

Looking into the residual inequality by quantile regression techniques is done by dividing the sample into very strict groups of people with the same characteristics and running quantile regressions within each group to see if there is considerable inequality. Although informative, it is very time-consuming, and the results are prone to be sensitive to the assignment of groups[5]. In that respect, JMP looks more practical.

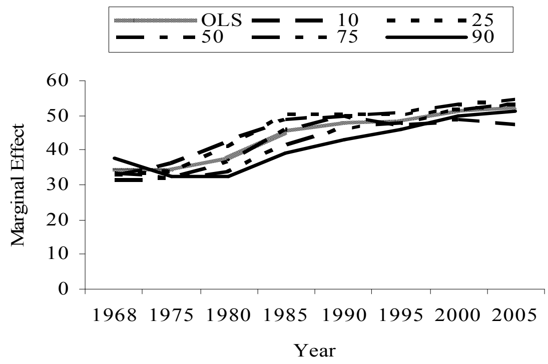
[5] See Autor *et al.* (2005) for an application.

**Figure 3**
Marginal Effects of Different Levels of Education

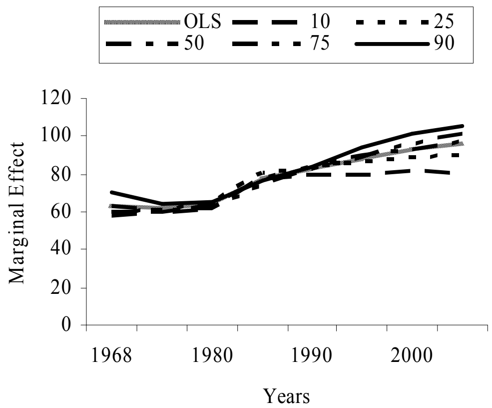
**A.High School**




**B.Some College**




**C.College**



## 5. Conclusion

We have analyzed wage inequality in the US for the past 39 years using CPS March data. The data shows that wage inequality increased sharply during the 1980s and then it slowed down at the beginning of 90s, increasing at a slower rate after that.

Obviously, both methods have their advantages in terms of the information they provide. Our analysis with JMP method helped us see the composition of wage inequality, confirming earlier findings by JMP (1993) that the "residual", or "within" inequality constitutes a large part of the increase in wage inequality starting from the 80s. We have seen that the effects of observable quantities and prices cancel each other out at times, the latter being larger. Our analysis also points that there might be an increase in the relative magnitude of the effect of observable quantities on the overall wage inequality after the turn of the century. Coupled with a limited magnitude of unobservable effect, this could mean a change in the composition of wage inequality, possibly meaning that the so-called "unobservable" skills have been absorbed by the educational system to some degree and/or can be learned consistently through work experience. Of course it is too early to draw strong conclusions from these results. We need to wait and see if this is just a few years' trend or a permanent change.

Our analysis using quantile regression confirmed the OLS findings (on which JMP is built) that college degree contributes to inequality more than other degrees do. This is both because having a college degree affects wages more than having one of the lower degrees and because there is dispersion among the people with college degrees. We also found that while the spread among the people with college degree was small prior to mid-90s and it started increasing after that, for people with lower degrees it was the other way around. The marginal effect of having either of the three degrees of education has been increasing at the upper tail and, to some degree, the upper half of the distribution, since the 70s.

The JMP decomposition enables us to identify the relative magnitude and shape of the effects of observable and unobservable skills and their prices on the increase or decrease of wage inequality. This is why it has been in use for the last decade and half despite the fact that many other methods of analysis have been proposed during the same period. However, as our analysis with the quantile regression estimates prove, one has to be careful about the shape of the distribution of wages when commenting on JMP results. The method itself depends on OLS estimates, which represent the conditional mean. When the distribution of wages is close to normal, one can comfortably make use of JMP results, since most of the data is centered around the mean and the tails are not strong. However, if the shape is different there might be problems. Since JMP decomposition is done for the changes in the difference of quantiles (90-10, 90-50, 50-10 in our case), such a distribution could limit the meaning of it. Thus, one should always check the quantile regression results before reaching any conclusions on the JMP decomposition technique results.

## Appendix

The annual wage and salary income entries in ASEC are given in top coded form. We impute the top coded values as 1.33 times the reported maximum value of the variable for that year. Then we deflate the annual wage and salary income to 1982 values, using the personal consumption expenditure deflator from National Income and Product Accounts (NIPA) which can be downloaded from the website of the Bureau of Economic Analysis (BEA, www.bea.gov/bea/dn1.htm). The natural logarithm of this value divided by the number of weeks worked during the reference year gives us average log weekly wage and salary income. Wage and salary income was reported as a single entry in CPS March data before 1987. However, primary and secondary job earnings started to be reported separately after this year. We imputed the top coded values with 1.33 of the top coding value separately before adding them up to find the total wage and salary income. One obvious problem is that top coded values show big fluctuations between years, especially for the secondary job earnings. Wage and salary income is available in six digits for the entire range of years. However, we have had some trouble with the weekly wages. Number of weeks worked during the reference year started to be reported as the exact number of weeks only from the 1975 data. It was a recoded variable before that, giving a value 1 to 7 representing a number of weeks in each group. We used an average of the number of weeks falling into each of these groups for 1975, 1976 and 1977 weighted by March Supplement Weights to recode the previous years' data into the actual number of weeks worked.

The educational attainment variable sees a few changes in its coding within the range of years we use. However, there is one major point of change: 1990. Before this year, the data is gathered from two separate variables, one that gives the highest grade of the school attended and another that states whether it is completed or not. If a person did not complete his/her last year of education, we simply assumed the years of education without that year. However, starting with 1991, the reporting of the years of education changed significantly. It became impossible to follow the exact years of education, with several years of education together in groups and reporting the same number for each group, but clearly stating if the person has a degree at any point. For the purposes of our analysis, we were still able to gather the needed information to form our 4 dummies as they were explained above.

## References

AUTOR, D.H., KATZ, L. F. and KEARNEY, M.S. (2005), "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists," NBER Working Paper no. 11627.

BLAU, F.D. and KAHN, L.M. (1994), "Rising Wage Inequality and the US Gender Gap", *The American Economic Review*, 84(2),23-8.

————(1996). "International Differences in Male Wage Inequality: Institutions versus Market Forces", *Journal of Political Economy*, 104(4), 791-837.

BRAINERD, E. (1998), "Winners and Losers in Russia's Economic Transition", *American Economic Review*, 88(5), 1094-1116.

BUCHINSKY, M. (1994), "Changes in the U.S. Wage Structure1963-1987: Application of Quantile Regression", *Econometrica*, 62(2), 405-58.

————(1998), "Recent Advances in Qantile Regression Models: A Practical Guide for Empirical Research", *Journal of Human Resources*, 33, 88-126.

JUHN, C., MURPHY, K. and PIERCE, B. (1991), "Accounting for the Slowdown in Black-White Wage Convergence", in *Workers and Their Wages,* ed. by M. Kosters, Washington D.C.: American Enterprise Institute Press, 107-43.

————(1993),"Wage Inequality and the Rise in Returns to Skill", *Journal of Political Econo*my, 101(3), 410-42.

KAHN, L.M. (1998), "Against the Wind: Bargaining Recentralisation and Wage Inequality in Norway 1987-91", *Economic Journal*, 108(448), 603-45.

KATZ, L.F. and AUTOR, D.H (1999), "Changes in the Wage Structure and Earnings Inequality", Handbook of Labor Economics, Vol.3A.

KIJIMA, Y. (2005), "Why did Wage Inequality Increase? Evidence from Urban India", *Journal of Development Economics*, 81(1), 97-117.

KOENKER, R. and BASSETT, G. (1978), "Regression Quantiles", *Econometrica*, 46, 33-50.

MACHADO, J.A.F. and MATA, J. (2001), "Earning Functions in Portugal 1982-84: Evidence from Quantile Regression", *Empirical Economics*, 26,115-34.

————(2005), "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression", *Empirical Economics*, 26:115-34.

MARTINS, P.S. and PEREIRA, P.T. (2004), "Does Education Reduce Wage Inequality? Quantile Regression Evidence From 16 Countries", *Labour Economics*, 11, 355-71.

*National Bureau of Economic Research* (2008), Available from http://www.nber. org/data/cps.html.

ORAZEM, P.F. and VODOPIVEC, M. (2000), "Male-Female Differences in Labor Market Outcomes During the Early Transition to Market: The Cases of Estonia and Slovenia", *Journal of Population Economics*, 13(2), 283-303.

SCHULTZ, T.P. and MWABU, G. (1998), "Labor Unions and the Distribution of Wages and Employment in South Africa", *Industrial and Labor Relations Review*, 51(4), 680-703.

## Özet

### Ücret ayrıştırmasında JMP ve kantil regresyon yöntemlerinin karşılaştırılması

Juhn, Murphy and Pierce (1993)'in ayrıştırma tekniği ve kantil regresyon, ücret eşitsizliği analizinin önemli araçlarından ikisidir. JMP tekniği, ücretlerdeki değişmeyi üç parçaya ayırma ve "artık eşitsizliği"ni kolaylıkla gösterme avantajına sahiptir. Buna karşın kantil regresyonun avantajı ise farklı kantillerdeki ücret dağılımının detaylı bir resmini gösterebilmesidir. Çalışmamızda her iki yöntemi de ABD İşgücü İstatistikleri Bürosu tarafından yayınlanan Mart Ayı Güncel Nüfus Anketi(CPS) verilerine uygulayarak 1967–2005 döneminde ABD'deki ücret eşitsizliğinde meydana gelen değişiklikleri inceliyoruz. Bu konuda hangi yöntemin daha faydalı sonuçlar ürettiğini görmek için sonuçlarını karşılaştırıyoruz. Sonuçta, JMP değerleri üzerinde yorum yapmadan önce kantil regresyon sonuçlarını kontrol etmek gerektiğini, çünkü eğer kantil regresyon katsayıları OLS regresyon katsayılarından çok farklıysa (yani ücret dağılımı normal dağılımdan uzaksa), iki yöntemin sonuçlarının oldukça farklılaştığı ve JMP'nin uygulanmasının problemli bir hale geldiğini görüyoruz.

*Anahtar kelimeler*: Ücret eşitsizliği, ABD, ücret ayrıştırması, kantil regresyon.

*JEL kodları*: J31, C14.

## Table 2

Comparison of Quantile Regression and JMP coefficients

***A.Returns to Education***

| | High School | | | | | | Some College | | | | | | College | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **OLS** | **10** | **25** | **50** | **75** | **90** | **OLS** | **10** | **25** | **50** | **75** | **90** | **OLS** | **10** | **25** | **50** | **75** | **90** |
| **1968** | 21.9 | 24.6 | 22.6 | 19.9 | 18.9 | 20.8 | 34.4 | 33.1 | 33.0 | 31.6 | 33.5 | 37.8 | 62.6 | 58.0 | 60.0 | 60.5 | 63.6 | 70.2 |
| | (0.4) | (0.8) | (0.5) | (0.4) | (0.5) | (0.6) | (0.5) | (1.0) | (0.7) | (0.5) | (0.6) | (0.8) | (0.5) | (1.1) | (0.7) | (0.5) | (0.6) | (0.8) |
| **1975** | 22.6 | 25.6 | 23.3 | 21.0 | 19.1 | 19.2 | 34.3 | 36.4 | 33.8 | 32.1 | 32.0 | 32.4 | 62.5 | 60.1 | 60.9 | 60.1 | 60.7 | 64.4 |
| | (0.4) | (1.0) | (0.7) | (0.6) | (0.7) | (0.9) | (0.5) | (1.3) | (0.9) | (0.7) | (0.8) | (1.1) | (0.5) | (1.3) | (0.8) | (0.7) | (0.8) | (1.0) |
| **1980** | 25.6 | 30.3 | 29.3 | 25.5 | 22.1 | 20.7 | 37.6 | 42.5 | 41.2 | 36.7 | 34.1 | 32.6 | 63.9 | 64.0 | 65.4 | 62.5 | 61.8 | 65.3 |
| | (0.4) | (0.9) | (0.6) | (0.5) | (0.6) | (0.7) | (0.5) | (1.0) | (0.7) | (0.6) | (0.7) | (0.8) | (0.5) | (1.0) | (0.7) | (0.6) | (0.7) | (0.8) |
| **1985** | 29.8 | 33.5 | 34.7 | 30.7 | 26.1 | 22.7 | 45.7 | 48.7 | 50.5 | 46.2 | 41.7 | 39.1 | 77.9 | 77.7 | 80.9 | 77.3 | 74.5 | 76.4 |
| | (0.5) | (1.1) | (0.7) | (0.7) | (0.7) | (1.0) | (0.6) | (1.3) | (0.8) | (0.7) | (0.8) | (1.1) | (0.6) | (1.3) | (0.8) | (0.7) | (0.8) | (1.1) |
| **1990** | 30.0 | 32.0 | 31.9 | 31.5 | 29.1 | 25.4 | 48.1 | 50.0 | 50.3 | 49.8 | 46.6 | 42.9 | 82.5 | 79.9 | 83.3 | 84.1 | 82.3 | 83.4 |
| | (0.6) | (1.2) | (0.9) | (0.7) | (0.7) | (1.1) | (0.6) | (1.4) | (0.9) | (0.7) | (0.8) | (1.2) | (0.6) | (1.4) | (0.9) | (0.7) | (0.8) | (1.1) |
| **1995** | 31.3 | 30.3 | 33.3 | 33.9 | 31.5 | 28.5 | 48.4 | 47.1 | 50.5 | 50.6 | 47.8 | 45.9 | 88.2 | 79.4 | 86.6 | 89.8 | 89.2 | 94.6 |
| | (0.7) | (1.4) | (0.9) | (0.8) | (0.8) | (1.3) | (0.7) | (1.4) | (1.0) | (0.9) | (0.9) | (1.3) | (0.7) | (1.5) | (1.0) | (0.9) | (0.9) | (1.3) |
| **2000** | 32.2 | 30.4 | 33.9 | 34.7 | 32.6 | 29.4 | 51.3 | 48.7 | 51.8 | 53.2 | 51.7 | 50.1 | 93.3 | 82.2 | 88.5 | 93.4 | 96.0 | 101.5 |
| | (0.7) | (1.2) | (0.9) | (0.8) | (0.9) | (1.3) | (0.7) | (1.2) | (0.9) | (0.8) | (0.9) | (1.4) | (0.7) | (1.3) | (1.0) | (0.8) | (0.9) | (1.4) |
| **2005** | 32.0 | 27.9 | 32.6 | 34.0 | 32.2 | 31.4 | 52.4 | 47.3 | 53.2 | 54.8 | 53.3 | 51.2 | 96.7 | 80.9 | 90.3 | 97.2 | 101.4 | 105.4 |
| | (0.7) | (1.5) | (1.0) | (0.9) | (1.0) | (1.6) | (0.7) | (1.5) | (1.1) | (0.9) | (1.1) | (1.7) | (0.7) | (1.6) | (1.1) | (0.9) | (1.1) | (1.6) |

***B.Impact Upon Dispersion***

| | High School | | | Some College | | | College | | |
|---|---|---|---|---|---|---|---|---|---|
| | **90-10** | **90-50** | **50-10** | **90-10** | **90-50** | **50-10** | **90-10** | **90-50** | **50-10** |
| **1968** | -3.8 | 0.9 | -4.7 | 4.7 | 6.2 | -1.5 | 12.2 | 9.7 | 2.5 |
| **1975** | -6.4 | -1.9 | -4.6 | -4.0 | 0.3 | -4.3 | 4.4 | 4.3 | 0.0 |
| **1980** | -9.6 | -4.8 | -4.8 | -9.9 | -4.0 | -5.8 | 1.3 | 2.8 | -1.5 |
| **1985** | -10.8 | -8.0 | -2.8 | -9.6 | -7.1 | -2.6 | -1.3 | -0.9 | -0.4 |
| **1990** | -6.6 | -6.1 | -0.5 | -7.1 | -6.9 | -0.2 | 3.5 | -0.7 | 4.2 |
| **1995** | -1.8 | -5.4 | 3.6 | -1.2 | -4.7 | 3.5 | 15.2 | 4.8 | 10.4 |
| **2000** | -1.0 | -5.2 | 4.2 | 1.4 | -3.1 | 4.5 | 19.4 | 8.1 | 11.3 |
| **2005** | 3.4 | -2.6 | 6.1 | 3.9 | -3.6 | 7.5 | 24.5 | 8.2 | 16.2 |

*Standard errors are reported in parentheses. All coefficients are multiplied by 100.*

**Table 3**

Comparison of Quantile Regression and JMP

***Changes in Dispersion(Conditional Wage Distribution)***

| | 75-68 | 80-75 | 85-80 | 90-85 | 95-90 | 00-95 | 05-00 |
|---|---|---|---|---|---|---|---|
| 90-10 | -0.031 | 0.096 | 0.083 | 0.031 | 0.075 | 0.054 | 0.018 |
| 90-50 | -0.014 | 0.045 | 0.039 | 0.016 | 0.047 | 0.044 | 0.002 |
| 50-10 | -0.017 | 0.050 | 0.044 | 0.015 | 0.028 | 0.011 | 0.016 |

***JMP(Total Change)***

| | 75-68 | 80-75 | 85-80 | 90-85 | 95-90 | 00-95 | 05-00 |
|---|---|---|---|---|---|---|---|
| 90th-10th | 0.029 | 0.042 | 0.172 | 0.022 | 0.094 | 0.057 | 0.042 |
| 90th-50th | 0.014 | 0.006 | 0.062 | 0.060 | 0.067 | 0.047 | 0.042 |
| 50th-10th | 0.015 | 0.036 | 0.110 | -0.038 | 0.027 | 0.010 | 0.001 |

**Table 4**

Descriptive Statistics

| ***Year*** | **1968** | | | **1990** | | | **2004** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **N** | **Mean** | **Std.** | **N** | **Mean** | **Std.** | **n** | **Mean** | **Std.** |
| Weekly Pay | 67,852 | 408.55 | 234.08 | 77,797 | 456.47 | 298.75 | 106,302 | 552.06 | 546.39 |
| ***Demographics*** | | | | | | | | | |
| Age | | 37.67 | 10.40 | | 36.86 | 9.86 | | 39.18 | 10.43 |
| White | | 0.91 | 0.29 | | 0.87 | 0.34 | | 0.83 | 0.37 |
| Married | | 0.95 | 0.22 | | 0.66 | 0.47 | | 0.62 | 0.49 |
| SMSA St. | | 0.70 | 0.46 | | 0.65 | 0.48 | | 0.71 | 0.46 |
| ***Education and Experience*** | | | | | | | | | |
| Less than HS | | 0.31 | 0.46 | | 0.11 | 0.32 | | 0.11 | 0.31 |
| High School | | 0.39 | 0.49 | | 0.39 | 0.49 | | 0.32 | 0.47 |
| Some Coll. | | 0.14 | 0.34 | | 0.22 | 0.42 | | 0.26 | 0.44 |
| College Grad. | | 0.16 | 0.37 | | 0.28 | 0.45 | | 0.31 | 0.46 |
| E<=10 | | 0.26 | 0.44 | | 0.28 | 0.45 | | 0.23 | 0.42 |
| 10<E<=20 | | 0.27 | 0.44 | | 0.36 | 0.48 | | 0.29 | 0.45 |
| 20<E<=30 | | 0.27 | 0.44 | | 0.24 | 0.42 | | 0.30 | 0.46 |
| 30<E<=40 | | 0.20 | 0.40 | | 0.13 | 0.34 | | 0.19 | 0.39 |
| ***Industry*** | | | | | | | | | |
| Agriculture | | 0.02 | 0.12 | | 0.02 | 0.13 | | 0.01 | 0.11 |
| Mining | | 0.01 | 0.11 | | 0.01 | 0.11 | | 0.01 | 0.09 |
| Constr. | | 0.08 | 0.28 | | 0.09 | 0.28 | | 0.12 | 0.32 |
| Manufacturing | | 0.38 | 0.49 | | 0.27 | 0.44 | | 0.18 | 0.39 |
| Tr.Com&P.U. | | 0.10 | 0.30 | | 0.11 | 0.31 | | 0.11 | 0.31 |
| Trade | | 0.15 | 0.36 | | 0.17 | 0.38 | | 0.15 | 0.36 |
| Finance&Serv. | | 0.18 | 0.38 | | 0.25 | 0.43 | | 0.35 | 0.48 |
| Government | | 0.08 | 0.27 | | 0.09 | 0.28 | | 0.06 | 0.24 |

3-year averages are given, centered on the specified year.